ESE DATA AND INFORMATION MANAGEMENT PLAN

Presentation to ESISS/ESSAAC
February 19, 2004

Mathew Schwaller

NASA HQ Code YF

mathew.r.schwaller@nasa.gov

Goals for This Presentation

- Present near and long-term strategy for DIMP review by ESSIS/ESSAAC and others

- Review the ESE Data and Information Management Plan (DIMP) motivations, themes, conceptual approach, outline and notional contents
  - Review comments received from ESISS on 2/17

- Solicit participation from ESISS/ESSAAC for more detailed review and comment to the DIMP annotated outline, plus participation in drafting the manuscript

DIMP Review Strategy

- **TODAY**
  - Request participation of ESISS/ESSAAC in review of the annotated outline and development of the DIMP (perhaps via a small Writing Team of interested participants)

NOTIONAL (subject to ESISS/ESSAAC agreement)

- **By end of March**
  - Full ESISS/ESSAAC review of the DIMP annotated outline (the outline circulated to ESISS prior to this meeting)
  - Writing Team will generate **revised annotated outline** (including roadmaps) based ESISS and ESSAAC comments
- **By end of May**
  - Writing Team generates **draft V0** of the DIMP
- **By end of June**
  - **Draft Version 1** of the DIMP for external review through the full ESISS & ESSAAC, and subsequently by the NRC

Section 1 Overview

- Section 1 **introduces the vision** for ESE data management a data and information management environment that…
  - **facilitates capture, processing, analysis and synthesis of NASA's Earth System Science Observations and associated research results, and that**
  - **provides timely, accurate and affordable access to the resulting knowledge and information.**

- Section 1 introduces 4 key challenges:
  1. Ever increasing Earth science data volume and user demand
  2. Completing the information cycle, i.e., including use of data & information in models and benchmarking in decision support systems
  3. Evolving Enterprise objectives in response to new discoveries and emerging national priorities
  4. Continuous and inevitable technological change
  5. *Growing user expectation of **robust web-based and grid-enabled services** (MA)*

Sections 2 through 6 in Summary

- Section 2 **describes each of the core ESE data management functions**, and the interrelationships among them in detail
- Section 3 **ties the ESE data management functions to science drivers** and to the set of Enterprise capabilities illustrated in the inner ring of Figure 1
- Section 4 **defines the current state of each capability and provides a "roadmap"** that charts the evolutionary steps needed to support the products and outcomes illustrated in the middle ring of Figure 1.
  - Section 4 also describes the role that Enterprise partners play as both recipients of—and contributors to—the ESE data management cycle (illustrated by the outer "customers" ring in Figure 1).
- Section 5 **describes the Enterprise policies for strategic investments** and implementation approaches as they apply to data management
- Section 6 **concludes with an approach for measuring success**: the evaluation criteria, assessment tools, and benchmarks that will be applied in the assessment of Enterprise performance with respect to Earth science data management.

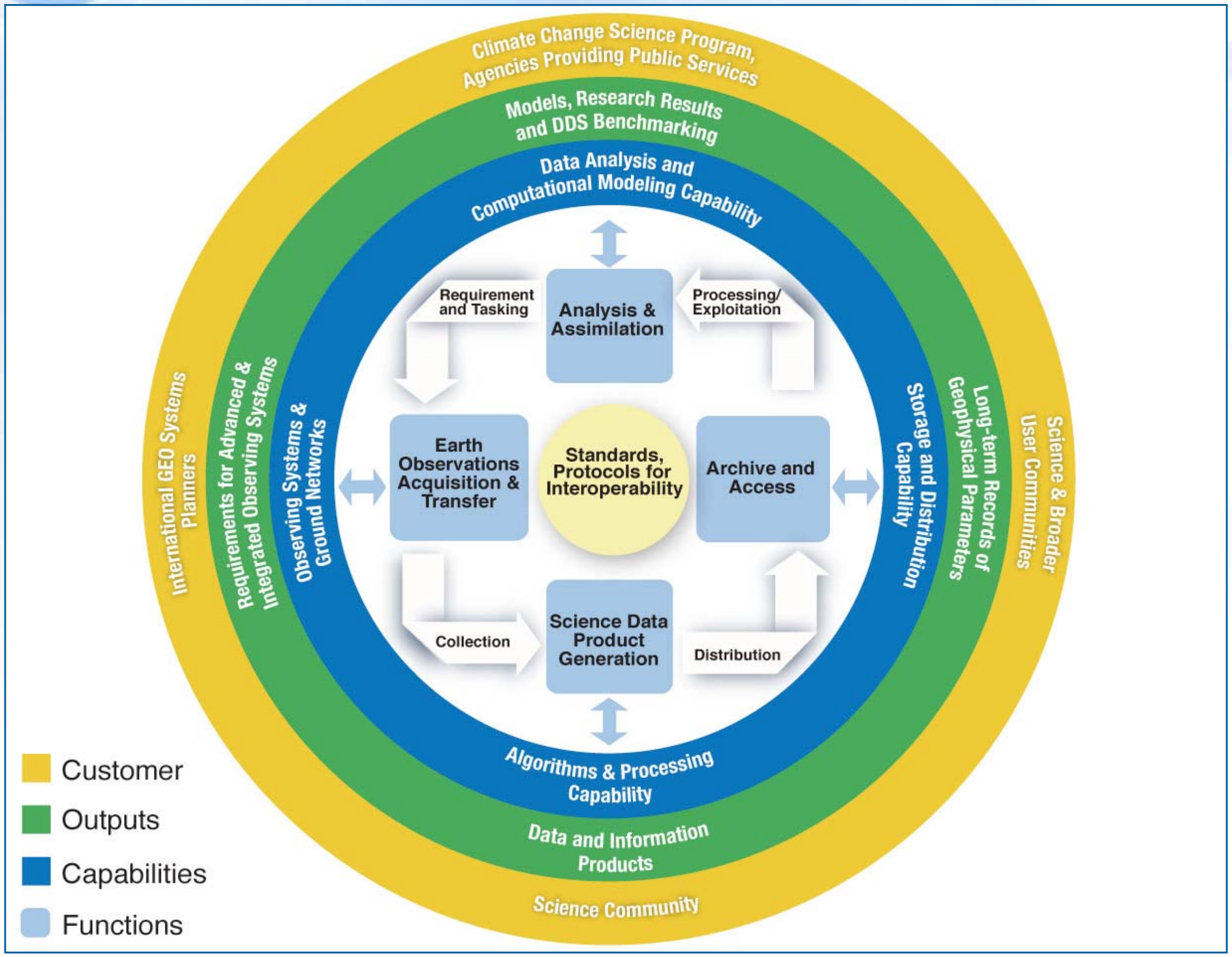
More on Section 1: Four Key Functions

- Earth observations acquisition and transfer for satellite data acquisition, operations ground networks and data transport.
- Science data product generation for retrieval of radiances and for generation of other geophysical data products from observational data to yield useful scientific data products and long-term measurement sets.
- Archive and access for making Enterprise data and information products available to a wide range of communities in, for example, science, applications, education and public policy.
- Analysis and assimilation for the integration of data products from NASA's ESE and other sources into Earth science models for understanding global and large-regional processes.

The plan should **reflect the vision of an integrated system, uniting sensors, archives, computing resources, etc as services** on a global web. It should not be based on the current paradigm which treats separately missions, product generation, archives, etc. (cf. The planetary nervous system paradigm.) ES

The plan should discuss an architecture and management structure that **allows new technology components to connect with existing systems** (ES)

Section 1: ESE Information Cycle



The interior of the figure identifies four major functions required of data & information systems in support of Earth Science

The four functions map to today's capabilities shown in the blue ring. Tracing the path out from the center as well as around the cycle helps set investment decisions in their larger context.

The diagram could be interpreted as **how NASA will circle the wagons to protect turf**… (JB)

Section 2 in More Detail

- Chapter 2 establishes linkage between the 4 ESE data management functions and current ESE institutional capabilities. It includes 4 sub-sections on functions
  - 2.1 Observations Acquisition and Transfer
  - 2.2 Science Data Product Generation
  - 2.3 Access and Archive
  - 2.4 Analysis and Assimilation

  Plus an additional sub-section on standards & protocols
  - 2.5 Making it Flow: Standards and Protocols for Interoperability

This section should build on the theme of "stability and innovation" perhaps mapping this theme into the 4 (5) functions

May be better to **define a set of mandatory principles** (core mission JB) that are needed, rather than focus on functions, **e.g. data stewardship** (TH, JB, RK, DE), **data security** (DE), **optimize use of computing resources** (RK), evolve-ability/**adaptability** (SG)

Section 3 in More Detail

- **3. CHARACTERISTICS OF AN EVOLVED DATA & INFORMATION MANAGEMENT ENVIRONMENT**

  3.1 A Foundation for the Future
  - How ESE will build on current capabilities

  3.2 Earth Science Research and Applications Drivers
  - Science research and applications drivers provide the motivation & rationale for ESE data management

  3.3 Characteristics of an Archetypal “Measurement System”
  - Defines the Earth-science research/applications-driven system characteristics needed to meet the Challenges of Section 1

  3.4 Implementation Approaches
  - Some illustative examples for implementation of the archetypal “measurement system”

Section 4 in More Detail

- **4. EVOLUTIONARY PATHWAYS: ACHIEVING A BALANCE OF STABILITY AND INNOVATION**
  - This section **defines a set of “roadmaps”** (1 for each of the functions described in Section 1) that **describe the evolution of these functions/capabilities over time**
  - This section also describes ESE data management **partnerships**

Notes from discussion on 2/17

Section 4 & 5 need revision

Section 4 needs to **emphasize the balance of stability & innovation** (DE)

a. Define which parts of EDE data systems require stability/stewardship (in other words, what **part of data systems need to be conservative, and which are ripe for introduction of new technologies?)**

b. **Move Section 5.4 into Section 4 to describe innovation/infusion**

c. What is the **methodology/process for introduction of new data management innovations**? Emphasize need for continual investment in prototypes (DE, PM)

d. **Additional goals** are: cost efficiency, maintenance of fundamental services (e.g., authentication, provenance, usability, stewardship)

Some ESISS **members questioned the utility of roadmaps as a methodology** for this report (DE)

Be sure that data roadmaps **trace to science roadmaps**

For “partnerships” **include private sector involvement** (e.g., s/w vendors (JD), OGC)

Sections 5 & 6 in Outline

- **5. MANAGING THE IMPLEMENTATION**
  - 5.1 HQ, Center, and Partner Roles
  - 5.2 The Competitive Process
  - 5.3 Implementation Path: Spiral Development and Next-Level Documentation
  - 5.4 Technology Investment and Infusion
- **6. PERFORMANCE ASSESSMENT**
  - 6.1 Goals and Objectives
  - 6.2 Evaluation Criteria
  - 6.3 Assessment Tools and Benchmarks

Section 5 needs revision as noted in the previous slide (or perhaps make Section 5.4 its own chapter RK)

Investment/infusion should: include COTS (JD), be influenced by user needs (MP), distributed computing (SG)

How will the heterogeneous environment be managed/governed? (JT)

Should engage interdisciplinary groups (GEWEX,…) in planning (SS)

Plan must allow for smooth response to disruptive change (PM)

Review Strategy *Revisited*

- TODAY
  - Request participation of ESISS/ESSAAC in review of the annotated outline and development of the DIMP (perhaps via a small Writing Team of interested participants)

NOTIONAL (subject to ESISS/ESSAAC agreement)

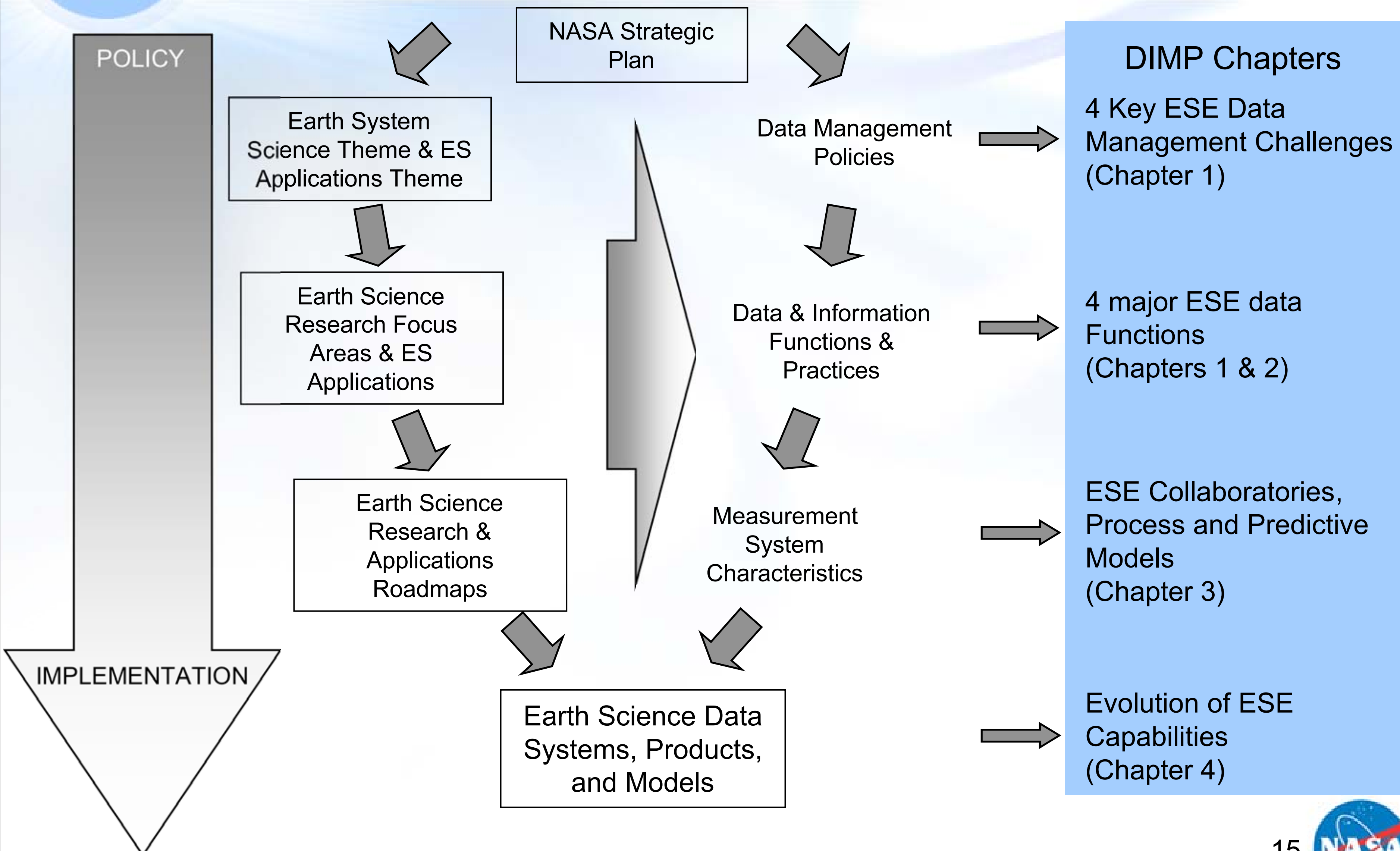
- By end of March
  - Full ESISS/ESSAAC review of the DIMP annotated outline (the outline circulated to ESISS prior to this meeting)
  - Writing Team will generate revised annotated outline (including roadmaps) based ESISS and ESSAAC comments
- By end of May
  - Writing Team generates draft V0 of the DIMP
- By end of June
  - Draft Version 1 of the DIMP for external review through the full ESISS & ESSAAC, and subsequently by the NRC

BACKUP CHARTS

DIMP Outline

NASA's ESE Data and Information Management Plan

How NASA/ESE Policies Map to the DIMP



Is this an adequate scope for the document?

OBSERVING SYSTEM MISSIONS & GROUND NETWORKS

TODAY'S CABABILITIES

- Satellite tracking ground stations for S-band and X-band with satellite direct broadcast X-band ground terminals.
- Geo-Stationary data relay satellites for ground-to-ground and space-to-ground data.
- Terrestrial optical fiber networks connecting major population centers using Internet Protocol data packet routing standards.
- Some capability for tasking satellite acquisition for defined science goals and observations of opportunity.
- Direct broadcast from some EOS space assets allows users to by-pass NASA infrastructure and acquire and process data as needed for specific user communities.

KNOWLEDGE STEPS

Space Network demand access services.

Software re-programmable receivers allows modifications and improvements of existing infrastructure.

Cost reduction in electronically steerable antenna arrays makes data transmission more secure.

Ka band space and ground systems improve transmission speeds.

Transmission protocols for interrupted service of data transport.

2003 → 2013

NOTIONAL EXAMPLE ONLY

10 YEAR OUTCOME

- Migration of telemetry downlink from 150 MBPS at X-band to 500 MBPS at Ka-Band, use of NPOESS Safety Net Ka-band network.
- Routine initiation of transient phenomena observations; MODIS detection triggers Hyperion observation.
- On-board data processing to provide quick-look data while enabled and coordinated space assets can achieve phenomena-triggered observations.
- Data fusion between instruments and observations are performed prior to relay to Earth-based systems.

Capturing and Transporting Data ...

DIMP Roadmap 1: Observing System Missions & Ground Networks

- NOTIONAL EXAMPLE ONLY
- Not intended for review today

- Other DIMP roadmaps include:
  - Algorithms & Processing
  - Storage & Access
  - Data Analysis, Assimilation & Modeling
  - A 5th roadmap on standards, protocols and interoperability is under consideration

Should Roadmaps be based on Functions or Capabilities?